EVALUATING THE ACCURACY OF SMALL AREA POPULATION ESTIMATES IN BRITAIN

Stephen Simpson, University of Manchester and City of Bradford Metropolitan Council
Ian Diamond, University of Southampton
David Lunn and Elizabeth Middleton, University of Manchester

Correspondence: Stephen Simpson, Centre for Census and Survey Research,
FESS-Dover Street, University of Manchester, M13 9PL, Manchester, UK.
email: stephen.simpson@manchester.ac.uk

NOTE TO CENSUS BUREAU: TABLE 2, AND FIGURES 1, 2 AND 3 MUST BE SCANNED IN FROM THE PAPER ORIGINAL SENT EARLIER. THANKYOU

ABSTRACT

Small area population estimates are provided for intercensal years by some local government authorities and commercial organisations, each using a different method; the accuracy achieved by these producers in 1991 is modelled here within a multilevel regression framework. Characteristics of the small area and of the method of estimation are included as explanatory variables. Results show that the nature of the areas to be estimated is strongly influential on the likelihood of achieving an accurate estimate. Independently of size and rate of change of population, areas with a concentration of poor, or black, or institutionalised residents are difficult to estimate accurately. In addition the method of estimation used is of great importance. Individual producers use the results of this evaluation to assess cost-effective ways of improving their methodology.

An account of the statistical modelling approach and its results form a major part of the paper. However other methodological issues encountered are addressed in order to assess the opportunity of using the 2000/1 round of censuses as a benchmark for evaluating small area estimates. These issues include: timely and relevant data collection; deriving a `true population' for fair comparison between different producers; choice of outputs that are relevant to policy applications of demographic estimates; dissemination and use of results. The *Estimating with Confidence* programme which undertakes this work in Britain was established by the local government sector with central government involvement, and is now funded by the Economic and Social Research Council.

KEYWORDS

Great Britain, Inaccuracy, Population estimates, Multilevel modelling, Small areas

1 INTRODUCTION

1.1 Our interest in accuracy

In this paper, small areas are usually between 1,000 and 30,000 population, and are sub-divisions of Local Government Districts. For the Districts, there is an accepted government method of annual component cohort estimation, which often acts as an independent constraint to which small area population estimates are made consistent. Our small areas may correspond best to incorporated places, minor civil divisions and small Counties in the USA, while Districts correspond most closely to larger USA Counties in size (Long 1993).

In Britain, interest in the accuracy of small area population estimates is expressed by:

- Producers of estimates, who are many and varied. They include over 100 local government organisations, who initiated this project and are most interested in improvements to their methods which have not before been quantitatively assessed.
- Government statisticians, who have committed themselves to producing population estimates for small areas, they are interested in knowing whether a single strategy with nationally available datasets can hope to be accurate enough to satisfy the critical eye of existing producers.
- Users in the media industry, whose audience ratings and therefore advertising revenue currently refer to two sets of conflicting population estimates with a resulting lack of confidence in the process.

An additional interest which is not strongly expressed, is the production of statistical confidence intervals around population estimates.

In the USA the evaluation of different methods, data sources, and data quality is the aim of recent Census Bureau work (Davis, 1994). A tendency for State organisations to depend on Federal work has caused concern for the quality of the resulting estimates (FSCPE 1995: 3). The Bureau of the Census state that they wish to integrate their variety of methods more efficiently, but without losing the flexibility of multiple methods and data sources (Long 1993: 14). These interests and concerns span the Atlantic.

We acknowledge that in a comprehensive evaluation issues other than accuracy are also important, including the political acceptance of results, and a firm delivery date of regular estimates; but we focus almost entirely on accuracy in this paper.

1.2 Concepts, aims and structure

We distinguish an estimation *strategy* from its components: data sets, consultation, and the use of a variety of estimation *methods*. A unified strategy that uses different datasets and methods on different small areas seems to be generally accepted in the US Bureau of the Census work, even if sometimes seen as unsatisfactory. This contrasts with the UK where a single method and dataset are seen to be necessary for equitable treatment of each area of the country in all official population estimates by central government.

The conclusions of our project hope to point to a multi-method, multi-data set strategy, where the best methods are used efficiently on each area estimated. Our analytical aim is to find which methods are usefully accurate in which types of area. This analytical aim requires a typology of methods, a typology of areas, an estimation of the true population to use as a gold standard for measuring the accuracy of estimates, a disentangling of the large area (District) estimates methodology from that of the small area methodology in which we are most interested, and measures of accuracy that reflect the public costs of getting the population size wrong.

We begin with a description of the dataset available to us (Section 2), and follow this with a typology of methods in use in Britain (Section 3) and a discussion of our solutions to some of the problems of evaluating accuracy of population estimates (Section 4). An examination of mean percentage absolute accuracy for subsets of areas shows that this approach has limited value for predicting the accuracy of different methods (Section 5). Then we provide an account of our statistical approach to evaluation (Section 6), and a summary of our preliminary results (Section 7). A discussion of the use of these results and of further work concludes the paper (Section 8).

2 THE ESTIMATING WITH CONFIDENCE DATA SET

Five elements make up our data set:

(a) The estimated 1991 population, defined as from a particular producer of estimates with defined methods not including use of the 1991 UK Census results, for a particular small area, for a specified age group. We have approximately fifty eight thousand such estimates in total, from 48 producers estimating two thousand separate areas. The structure of this estimates dataset is complex, as the same area may be estimated by one, two or three different producers, some without and some with an age disaggregation (using several different age disaggregations!).

(b) For each estimate, the equivalent `gold standard' population (based on results from the 1991 census), the true population against which the accuracy of the estimate is to be evaluated. The difference between (a) and (b) is the inaccuracy of the estimate.

(c) For each estimate, the equivalent population in 1981 from the 1981 census.

(d) Characteristics of each small area, based on the 1991 census results.

(e) Characteristics of the method used by each producer.

The merging of these five data sets on a consistent basis was no easy task. It involved creating standard census output for non-standard areas by a variety of strategies, coping with changes of area boundary between 1981 and 1991, understanding the population universe targeted by each producer, and dividing producer datasets such that same combination of methods was used for all small areas of one producer. That work is complete as of March 1996 (Lunn et al. 1996).

3 ESTIMATION METHODS

3.1 The extent of small area population estimation in Britain

This section draws on the project's review of methods which is more fully reported in Simpson et al. (1996). It was necessary to characterise the methods of estimation before comparing their accuracy. A survey of local authorities' sources of estimates of the current population of small areas (before the 1991 census results had been released) found that fully 15% of respondents relied on the output from the previous census, held in 1981, without any update. Thirty per cent of responding authorities - 101 separate organisations - had made at least one update to local population estimates since 1981, most of them on a regular basis. Half of these estimated the number of households in each area, while one third estimated an age structure in at least five year age group detail. These 101 authorities were mainly the Metropolitan, County and Regional authorities responsible for larger regions within Britain, who were therefore estimating a large number of areas each and covering between them the majority of places within Britain.

In the USA, only the total population and total households are regularly estimated for areas as small as ours. This is a reflection of the lack of suitable data for reliable detail (Long 1993).

3.2 Broad types of estimation method

Forty eight local authority producers of population estimates subsequently provided the project with both data and details of their estimation methods. Table 1 summarises their methods and data sets.

The methods are listed in order of increasing complexity and expense, starting with apportionment of an estimate for a larger area population, invariably the District containing the small areas. *Apportionment* requires only an indicator of the current population in each small area, no previous data, and assumes that the relationship of the indicator to the true population is the same in each small area. The two `change' methods update the previous census results for each small area; the *ratio change* method calculates and applies a rate of change derived from measure(s) of population stock at the time of the previous census and currently; the *additive change* approach uses indicators of additional population, usually involving changes to housing stock.

Cohort survival methods involve at least the ageing of the base population at the time of the previous census and estimation of young children from births since the previous census. Finally, a significant number of local authorities attach a short demographic *census* to the annual canvass for local electoral registration.

When the measure of change is from a dwellings database, the ratio method corresponds to the `Housing Unit' method common in the USA for small areas (Smith, 1986). While the other methods are simple in design and computation, they do not have a direct relationship to methods used in other countries. This is because, as in other countries, they are based on the data sets available to those producing the population estimates. In Britain, two main sources of data for small areas have been used: the electorate, which in Britain includes 93% of eligible adults, and the monitoring of regulated changes to dwellings: new and demolished dwellings, and changes of use.

3.3 Combinations of methods and datasets

The forty eight producers are classified in Table 1 according the most complex of the five listed methods which they use. As mentioned earlier, producers often combine more than one method. For example the total population may be set according to an change method and the age structure according to a cohort or local census approach. Almost all producers constrain an initial estimate to add across small areas to some independent estimate of the District population, often the government estimate.

The interest in a nationally-applicable method, and the falling quality of the electorate size as an indicator of population size, has led to the investigation of patient records of the National Health

Service as a demographic source. These are likely to be an important ingredient of demographic work in Britain at all area levels during the next decade, but are seldom in practical use at present.

As part of the database describing the method of each producer, characteristics that may be of particular use in the evaluation of accuracy are:
- Main method used to compute the total population.
- Main method used to compute an age structure (if any).
- Use of the electoral register as a major determinant of the adult population, and/or the total population.
- Use of dwellings statistics as a major determinant of the total population.
- Whether any data have been used which are not available consistently on a national basis.
- Target date (mainly 1991, but occasionally 1990 or 1992).
- Number of years for which the estimate was rolled forward as a projection (a small number of producers estimated for a 1990, 1989 or 1988 population, and then projected it forward one, two or three years).

We also have some limited data referring to the resources expended in using each estimation procedure, which are greatest for the local census.

4 METHODOLOGICAL ISSUES REVIEWED: A TRUE POPULATION, LARGE AREA ESTIMATES AS CONSTRAINTS, AND MEASURES OF ACCURACY

In addition to the difficulty of putting together a dataset from many sources, and the need for a typology of methods, discussed above, those who wish to evaluate the accuracy of different methods have to solve several other analytical issues. The statistical modelling of accuracy is discussed in sections 6 and 7. Here we address (a) the nature of a true population, (b) the influence of inaccuracy of large area estimates on small area estimation methods, and (c) appropriate measures of accuracy.

4.1 A true population count

Most users of population estimates assume that they are provided with an estimate of a population which includes all residents within a defined geographical boundary, perhaps with some qualification as to the inclusion of those in communal establishments (group quarters in the USA) and disaggregation by specified age groups. Measuring the accuracy of a population estimate therefore involves estimating the true population by some better standard, and is thus usually only attempted for years in which a population census has been taken. Members of the Estimating with Confidence team were closely involved in the research in Britain to estimate the size and geographical distribution of the 1991 Census undercount, which was both greater and less well monitored than in previous censuses in Britain.

The project undertook technical development, testing and consultation to achieve a plausible, consistent and accepted method of estimating census undercount in each small area of Britain and then a good post-census estimate of the full population in each small area. The method is summarised in Appendix 1, and more detail is given in Simpson et al., 1995. This has been used as the gold standard against which to measure the accuracy of the producers' pre-census estimates of population size. There is inevitably some measurement error attached to this gold standard of which we must be conscious during the analytical stage of the project, particularly for young adult males and for inner city areas. Nonetheless, the acceptance of a single set of accepted census adjustments has been a major achievement. We believe that the raw census count is inadequate as a standard for asessment of estimates of the full population.

4.2 The influence of large-area constraints

Since most producers of small area populations constrain their estimates to be consistent with estimates for Districts derived independently (see section 3), the inaccuracy of each estimate provided to the project is in part due to the estimation methods used for the District population. In order to focus on the accuracy of the methods used to estimate the small areas, the analytical results (in Section 7) have used the producers estimates constrained to add to the District population. We are in effect evaluating the accuracy of producers' estimates of the distribution of population within each District. This also corresponds to a major use of population estimates: distribution of resources within a larger area.

4.3 Measuring inaccuracy

Finally, how should inaccuracy be measured and summarised for a set of population estimates? Many measures have been proposed. Seven were used throughout recent work in the USA (Davis, 1994), five of which were highly correlated and not discussed separately. We take the view that measures of accuracy should allow the users `loss function' to be assessed, ie they should reflect the penalties incurred by users from different levels of inaccuracy. However, since there are so many varied uses of population estimates this does not point to a single measure of inaccuracy.

In the next section we present summary measures of accuracy over a set of population estimates, using the weighted Mean Absolute Percentage Error (MAPE). The absolute percentage error for each area is weighted by its true population. This in effect counts the absolute difference between estimate and true population across all areas and expresses that total inaccuracy as a percentage of the total population in all areas. It can be usefully interpreted as (twice the) percentage misallocation. The percentage of large inaccuracies (10% or more) is also referred to.

In Section 7 our analytical approach proceeds to examine the individual population estimates. There we retain the absolute percentage error as our measure of inaccuracy, but have to transform it logarithmically in order to achieve a normal distribution for statistical testing procedures. A multinomial approach is being developed to also explain the occurrence of large absolute errors.

5 MEAN INACCURACY

5.1 Tabulations show some effect of different types of area and methods on innaccuracy

Table 2 presents the mean inaccuracy (MAPE as described above) of small area population estimates, within a variety of categories describing the type of area and the method of estimation used. The types of area have been chosen to correspond with the size and growth categories presented for USA County estimates by Sam Davis (1994).

The mean inaccuracy for all the estimates of total population is 3.8%. As expected, and reflected also in USA evaluations, the estimates are less accurate for small areas, and particularly for small areas that have experienced population change during the decade since the previous census (of 5% or more), where the mean inaccuracy rises to 5.4% and over 1 in 5 estimates are at least 10% inaccurate.

It is more difficult to estimate the age structure of a small area than its total population. It is most difficult to estimate the number of young adults aged 15-24, the most mobile and diversely housed group (MAPE of 14.8%). It is also relatively difficult to estimate the number of children and the number of elderly (MAPEs of 10.6% and 12.4% respectively). This is important for those concerned with providing public schooling and care.

The estimates provided by different methods, as broadly categorised in section 3, show some (generally smaller) differences in accuracy achieved. Cohort survival methods gained a lower mean inaccuracy while local censuses achieved fewer large inaccuracies.

But is the accuracy of some methods simply due to their having been used in areas that are easier to estimate: large and slow-changing areas for example?

Table 2 cross-classifies the method of estimation with the characteristics of the area estimated, and does suggest that the local census method is coping relatively well with difficult as well as with easily estimated areas. However, the number of areas in each category is now rather small and one should beware generalising too much from comparison of pairs of cells in Table 2.

5.2 Beyond tabulating mean inaccuracy: modelling and experimentation

In fact there are several limitations of comparing the mean accuracy of sets of estimates from different areas and producers in the way presented by Table 2, in spite of its immediate use in painting clear broad-brush summary results.

First, the number of cells and therefore comparisons to make is too large to take in visually. Second, as already observed the number of areas has already become small in some of the comparisons and demands consideration of sampling error. Third, in spite of its unwieldy size the characteristics of areas and methods shown in Table 2 are still insufficient to be sure that the relative inaccuracies shown are not due to something that has not been measured in the table. For example, it may be that the rural-city dimension, or the siting of communal establishments within an area, would explain the differences apparently due to different methods. Alternatively, more detailed descriptors of the method and datasets used to make each estimate may account for the differences between the broad method types.

These limitations suggest that we need to look more closely at the differences between individual producers and at the same time take more account of the characteristics of the areas estimated. A different approach is required for this.

The project is proceeding in parallel with two new approaches. Firstly, an experimental approach has collated data for a limited number of small areas in Britain whose boundaries have not changed between 1981 and 1991. Estimation routines have been programmed to provide 1991 population estimates by each of the main methods of estimation, with several choices of data set and options for adjusting the 1981 census, inclusion of information about communal establishments, and constraint to independent estimates for the containing District. Over 1,000 different methods are available from these routines.

Because each method from this experimental approach will be applied to the same set of areas, summary measures of inaccuracy such as the MAPE will highlight real differences between methods, at least as applied to the experimental set of small areas. This approach seems to parallel that promised in further work by the US Bureau of the Census, who also found that comparison of groups of different counties where different methods were applied was difficult to interpret (Davis, 1994: 14).

In this paper however, we maintain our dataset of real estimates produced for different areas by different methods, and attempt to overcome the difficulties outlined through a statistical analysis of the individual inaccuracies of each estimate, using multilevel regression models. The approach is outlined in Section 6, and results given in Section 7.

6 MULTILEVEL REGRESSION MODELLING OF THE INACCURACY OF INDIVIDUAL POPULATION ESTIMATES

The *percentage inaccuracy*, the difference between the population estimate made by a producer and the true population, expressed as a percentage of the latter, shows the extent that the population estimate is above or below the true population. The distribution of this variable is displayed in Figure 1 and gives some immediately useful impressions. The majority of estimates are within 5% of the true population, but a significant number contain errors considerably in excess of 5% and these are of some significance to users of the estimates. The percentage inaccuracy centres on 0, implying that there is no general tendency to over or under-estimate local populations.

However, most influences on the inaccuracy of a population estimate concern the reliability of data sources and methods. They affect the size of the percentage inaccuracy, rather than its direction, and it is this that is of most practical importance. For example, larger populations are easier to estimate accurately but there is no evidence of any systematic bias towards over or under estimates. The dependent variable used in this paper is therefore the *absolute percentage inaccuracy*, transformed by taking logarithms to maintain approximate normality in the residuals of the model estimated. One is added to the absolute percentage inaccuracy so that the logarithm is defined even for population estimates that had zero error. The distribution of the dependent variable, $\log_e$(absolute percentage inaccuracy + 1) is shown in Figure 2.

To identify the characteristics of small areas which most influence levels of inaccuracy one requires some kind of regression framework. Here the data have a hierarchical structure with two levels - each small area, typically a ward (level 2), has one or more estimates (level 1), each made by a different producer. Such hierarchical or multilevel structures can now be analysed straightforwardly using multilevel models (Prosser et al, 1991).

A multilevel model with random effects for individual estimates and for small areas can be specified as follows:

$$Y_{ij} = a + \underset{\sim}{b}' \underset{\sim}{X}_{ij} + \underset{\sim}{\gamma}' \underset{\sim}{Z}_j + \lambda_j + e_{ij} \qquad (1)$$

where Y_{ij} is the measure of inaccuracy, $\log_e$(absolute percentage inaccuracy + 1), the X_{ij} are characteristics of the estimate i, varying within the set of estimates for a small area j, the Z_j are characteristics of the small area, λ_j is a residual random effect of unmeasured small area characteristics, and e_{ij} is random error in the estimates remaining for individual estimates.

7 RESULTS

7.1 Variation in inaccuracy explained

In this paper, results are given for estimates of the total all-age population of small areas. The multilevel model specified in the previous section was estimated using the ML3 program (Prosser et al., 1991). The socioeconomic, demographic and producer variables included are summarised in Table 3. A parsimonious model was selected using a combination of forward selection and backwards elimination. The results are presented in Tables 4 and 5. Table 4 presents the coefficients and standard errors for the parsimonious model including variables to reflect both the method used to make the projections and the demographic characteristics of the area. It is clear that there is a wide range of variables which influence accuracy.

An easier way of interpreting these results is given in Table 5, which presents in its first row the inaccuracy predicted from the model when all explanatory variables are held at their mean value. Subsequent rows of Table 5 give the inaccuracy predicted when each explanatory variable takes an extreme high or low value while all other variables in the model are held at their mean value (for categorical variables, the proportions in each category for the whole dataset were used). For the interaction, the inaccuracy predicted is given for each combination of extreme and mean values of the two variables concerned.

As expected, smaller areas are more likely to incur estimation errors, and this is particularly so for small areas which have seen considerable changes in population. Thus the model predicts that while the inaccuracy for an estimate of an area with mean characteristics is 3.1%, it is expected that this will increase to 5.5% in areas of only 1,000 population which have seen a change of 50% in their population since 1981. No predicted value is shown for an area with large population and large percentage population change, as this combination did not appear in the dataset and such an area would rarely exist in practice.

Very importantly, the expense of a local survey or census of population pays off, to the extent of gaining an expected extra 0.8% accuracy in each small area over other methods applied in similar areas. However, an equally important result is the absence in Tables 4 and 5 of a variable which indicates whether a producer made use of ancillary local data, as this variable was not significant in the model. Clearly this result needs further investigation, but it suggests that some methods of using data that are available for all areas of Britain have been relatively successful.

The socioeconomic and demographic characteristics which influence inaccuracy are exactly as one might expect. Areas with a high proportion of armed forces, institutional populations, and students resident in term time are relatively hard to estimate. These characteristics reflect mobile populations. On the other hand the remaining two significant variables, the proportions of unemployed and of black and asian residents, are proxies for urban and socially disadvantaged areas where for example the prevalence of multi-occupier households may make estimation difficult.

7.2 Unexpected variation in inaccuracy: is it under the control of producers?

There is a relatively large unexplained population estimate variation (level 1) with a confidence interval over twelve percentage points wide. This demonstrates the need for further work on the specific characteristics of the methods used by different producers. For example, a local census may be undertaken with different levels of staffing and expense, achieving different response rates and levels of accuracy. At the second level there is also unexplained variation, with a confidence interval which is around seven percentage points wide. This points to there being characteristics of local areas not measured in this model, for example the level of migration, but which make small areas within a District relatively hard or easy to estimate.

Figure 3 addresses the variation of inaccuracy between estimates in a way that illuminates the extent to which improvement is under the control of producers, a key interest at the outset of this paper. Figure 3a shows the accuracy predicted from each of the 29 producers, given the characteristics of their local areas and their choice of estimation method. For the model of Table 4, the predicted inaccuracy of each estimate has first been exponentiated and 1 subtracted. The mean and standard deviation across all areas of a producer, shown in Figure 3a, can then be interpreted directly as the predicted distribution of absolute percentage inaccuracy for that producer.

In the graph, producers have been grouped by their estimation method. For a particular method the characteristics of the small areas estimated clearly affect the likelihood that a producer will provide an accurate estimate. For example some producers using the simplest method, apportionment, can expect half the error of other producers using the same general method. This is because their areas are much larger and therefore easier to estimate accurately, or have fewer concentrations of those populations that Table 4 indicates give rise to larger inaccuracies, such as students, the unemployed and armed forces. The standard deviation for a producer reflects the variation in types of area that the producer has estimated. The first producer clearly has a uniformly easy set of areas to estimate, the fifth a much more challenging variety of areas, and the ninth a more uniformly difficult set of areas to estimate.

An important issue is the extent to which there remains variability between producers after allowing for the broad type of method used and the socioeconomic and demographic characteristics of the areas. Figure 3b shows approximate confidence intervals constructed for each producer's mean residual inaccuracy to test for differences in inaccuracy between producers. As before the residuals are calculated from the exponentiated modelled inaccuracy. Thus they are not the residuals direct from the model, but are the differences between the observed absolute percentage inaccuracy and its value predicted from the model. These differences are approximately normal; the confidence interval around each mean residual has been scaled down using the procedure proposed by Goldstein and Healy (1995) to construct simultaneous confidence intervals to test for differences between any pair of producers. The mean residuals for two producers are significantly different if the intervals do not overlap in Figure 3b.

Figure 3b has producers displayed in the same order as in Figure 3a. Immediately it can be seen that, within a particular method, there is considerable variation between producers in the inaccuracy of their estimates. For example, producers 2 and 3 both undertook a local census, for areas which Figure 3a shows are of similar difficulty of estimation, but Figure 3b shows that producer 2 incurred considerably more inaccuracy than producer 3. Such differences between producers displayed in Figure 3b may reflect effort put into making the estimates, variations in the way a particular method is operationalised, or some other unobserved factor.

Overall, Figure 3 suggests that the unexplained differentials between producers' inaccuracy, that may be due to factors under their control, are of the same order as, and rather larger than, the differentials explained by area characteristics and therefore outside the producers' control. This result is reason to pursue further improvements in local demographic estimation.

8 DISCUSSION

This paper has shown that although population estimates in England and Wales have been subject to a degree of inaccuracy, around 80% have been within five per cent of the true values, even ten years after the previous census. However the key results are those with regard to type of method used as it is these that will be of most relevance to producers who must decide which method to use to make their estimates.

To undertake meaningful comparisons, a national standard population for local areas in mid-1991 has been developed, albeit with some remaining measurement error that has not been explicitly modelled in this paper. Despite this drawback, there are a number of practical conclusions. First, the gradient of the reduction of inaccuracy with size of area is reasonably steep. This suggests that small areas are likely to be particularly difficult to estimate and that *total* population size in large areas may be estimated with reasonable accuracy. Second, although it is widely accepted that local censuses are expensive it does appear that they do have some benefit in improved accuracy. Third, it is not clear whether local population indicators do have a substantial gain over indicators which are available nationally. It may be the case that in areas which are particularly difficult to estimate, the benefits of local knowledge may become more apparent.

The identification of producers who appear to have been particularly successful is intended to lead to descriptions of good practice that may be helpful to others. Further analysis of the dataset will show the importance of the size and type of the producer's organisation, and of the effort expended on using particular local data sources. Of particular importance will be the need to consider different age groups for which a smaller dataset has been collated from producers. A further important focus will be an experimental approach. Here, different estimates will be made for the same areas, using each of several cohort survival, local census and apportionment methods and a full cost-benefit analysis will then be possible.

The next opportunity for assessing the accuracy of local population estimates will be the 2001 census. Our work suggests that a reliable measure of census undercount in small areas is extremely important to such assessments. It also warns against underestimating the work involved in matching census data from two censuses to population estimates provided by producers for a variety of spatial geographies. The project benefitted from the close involvement of producers over a period of years in the development stage of the project, which went some way to overcoming the difficulties of non-standard documentation of the geographies and methodologies lying behind each estimate.

The results of this and other studies also clarify that a method will work better in some areas than in others. What equity is there in maintaining a single method in all areas? Such even-handedness may result in costly work in areas that do not merit it, and lack of focus on the difficulties of areas that do merit more attention. The aim could rather be to maintain zero bias in all areas, and to improve the accuracy wherever possible. A single *strategy* may use measured characteristics which are known to be associated with the efficiency of different methods, in order to apply the appropriate method to each area and gain on overall accuracy in this way. We cannot yet provide the algorithm for such a strategy, but our studies go some way to showing its feasibility.

Acknowledgements

This research was supported by the Analysis of Large and Complex Datasets programme, funded by the Economic and Social Research Council (UK) award no. H519255028.

References:

Davis, Sam T (1994) Evaluation of postcensal county estimates for the 1980s. Population Estimates and projections technical working paper series no. 5. US Bureau of the Census: Washington.

Federal State Cooperative Program for Population Estimates (1995) Minutes of FSCPE Spring meeting, San Francisco April 3-4 1995. US Bureau of the Census: Washington.

Goldstein, H., Healy, M.J.R. (1995) The graphical presentation of a collection of means. *Journal of the Royal Statistical Society Series A*, **158**(1), 175-177.

Long, John F (1993) Postcensal population estimates: states, counties, and places. Chapter 4 in 'Indirect estimates in federal programs', Statistical policy working papers 21, US Presidents Office: Washington.

Lunn, David, Stephen Simpson, Ian Diamond, and Liz Middleton (1996) Problems of data merging census and demographic datasets. Estimating with Confidence Working Paper 12. University of Southampton Department of Social Statistics: Southampton.

Prosser, R., Rasbash, J., Goldstein, H. (1991) ML3: Software for Three-level Analysis, Users Guide. London: Institute of Education.

Simpson, Stephen, Liz Middleton, Ian Diamond and David Lunn (1996) Small area population estimates: a review of methods used in Britain in the 1990s. Estimating with Confidence Working Paper 11. University of Southampton Department of Social Statistics: Southampton.

Simpson, Stephen, Rachel Tye and Ian Diamond (1995) What was the real population of local areas in mid-1991? Estimating with Confidence Working Paper 10. University of Southampton Department of Social Statistics: Southampton.

Smith, Stanley (1986) A review and evaluation of the housing unit method of population estimation. Journal of the Am Stat Soc **81**: 287-296.

Appendix 1 Estimates of under-enumeration and mid-1991 population: small areas

Estimates of census under-enumeration and mid-1991 population are now available for small areas from the *Estimating with Confidence* project.

When OPCS in England and Wales, and GRO(S) in Scotland, provide population estimates in census year for local authority areas, the census figures are adjusted in three ways: they are moved forward from census day in April to 30 June, student numbers are transferred from home address to term-time address, and an allowance is made for incomplete census coverage.

The *Estimating with Confidence* project funded by the Economic and Social Research Council has undertaken similar adjustments for each ward of England and Wales, and each postal sector in Scotland. The adjustments are sensitive to local characteristics, but are consistent at each age and sex with the government estimates. The small area adjustments and the population estimates add up to the government local authority equivalents.

The new estimates are based on research by the project, based at Manchester and Southampton universities. Wide consultation during 1994 gave most support to two indicators of local census coverage among nine candidates which were tested. In the absence of hard evidence on local census coverage, these two were considered most plausibly associated with the distribution of census undercoverage within a local authority area, and when used in combination gave the most acceptable estimates of census undercoverage. They were (i) the number of male unemployed aged 20-34 and (ii) the number of persons imputed in census returns.

The results show ward undercoverage, excluding that related to armed forces, ranging from under 1% and up to 10% for the total population of a ward, and up to 50% for men in their twenties, about twice the range estimated for local authority Districts. The adjustments for students and for armed forces give more extreme adjustments, although to relatively few areas.

The estimates are available as a base for updating local population estimates in the 1990s, and as an indicator in their own right for resource distribution at a local level. While these estimates are subject to an unknown degree of error as with all population estimates, their advantage is consistency across all areas of Britain and their derivation with a method which takes into account the characteristics of each local area. OPCS have used the estimates to derive a mid-1991 population estimate for those new unitary local authorities in England that do not coincide with previous District boundaries.

A description of the methodology behind the estimates (Working Paper 10) and the estimates themselves, are available from Kathy Hooper, *Estimating with Confidence*, Department of Social Statistics, University of Southampton, SO17 1BJ, UK. Mid-1991 population estimates for Census Enumeration Districts, areas of Scotland, and other areas may be prepared on request. For academic research in Britain, the estimates are available from the MIDAS online datasets service at the University of Manchester, UK.

Table 1: Main estimation methods in Great Britain: local authority type, data sets

	Total no. of authorities	Counties and county districts in England and Wales	Metropolitan Districts and joint research bodies in England	Regions and Districts in Scotland	Indicator of current population	Indicator of change in population
Apportionment	13	8	3	2	Electorate only, 8 Electorate and dwellings, 3 Electorate and others, 2	
Ratio change	16	7	8	1		Electorate only, 7 Electorate and dwellings, 4 Dwellings only, 1
Additive change	8	2	3	3		Dwellings only, 7 Dwellings and other, 1
Cohort survival	4	2	2	0		Migration from dwellings change, 3 Other, 1
Local census	5	3	2	0	Electoral registration enhancement of all households,5	
Other	2	1	1	0	Various age-related indicators summed without adjustment	

Source: 48 producers included in Estimating with Confidence project analyses
Note: 'Dwellings' and 'Dwellings only' implies the use of a property or other register with information about the number of (or change to the number of) dwellings or households, and usually a persons-per-household ratio.

Table 2: Mean inaccuracy (weighted MAPE) of population estimates, Great Britain

		All areas			Areas with less than 5% growth 1981-1991			Areas with 5% or greater growth 1981-1991		
		All sizes	<2,500	2,500+	All sizes	<2,500	2,500+	All sizes	<2,500	2,500+
ALL METHODS	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									
APPORTIONMENT	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									
RATIO CHANGE	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									
ADDITIVE CHANGE	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									
COHORT SURVIVAL	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									
LOCAL CENSUS	All ages									
	0-14									
	15-24									
	25-44									
	45-64									
	65+									

Key to entries in each box: Left in large writing: weighted MAPE.
Top right: number of estimates. Bottom left: percentage of these estimates which were in error by more than 10%.
Top right corner: if shaded, all estimates were from one producer.
'All ages' refers to the accuracy of estimates of the total population, which were provided by all producers. the age structure was provided by a minority of producers.

Table 3: Summary statistics for variables to be included as explanatory variables in multilevel modelling

	Mean	Std. Dev.	Min	Max
Response Variable				
Log_e(Absolute percentage inaccuracy + 1)	1.41	0.72	0.00	5.56
Small area characteristics				
Absolute percentage change in population 1981-1991	9.48	21.46	0.00	804.95
Population Size, mid-1991	7,892	4,241	253	31,930
Black and Asian residents (%)	9.91	13.41	0.00	90.49
Armed Forces residents (%)	0.06	0.25	0.00	4.61
Institution residents (%)	1.60	2.55	0.00	28.27
Economically active residents (%)	50.31	4.74	28.58	67.77
Student term-time residents (%)	2.22	3.13	0.08	55.64
Unemployed residents (%)	9.57	5.67	1.37	44.11
Estimate characteristics	Number of estimates	Number of producers		
Overall type of method used				
Local census	753	5		
Cohort survival	1996	7		
Apportionment and ratio	2807	17		
Data sources				
Some only locally available	3573	25		
All nationally available	1983	4		

Note: The small area characteristics are taken as a percentage of all 1991 census residents, except for the variable unemployed residents which is taken as a percentage of economically active residents.

Table 4: Estimates and standard errors of a multilevel model to predict Log_e(absolute percentage inaccuracy +1)

Variable	Estimate	Standard Error
Constant	1.37	0.03
Absolute population change 1981-1991 (%)(APC)	0.01	.0015
Method used is Local Census	-0.22	0.03
Population size, mid-1991 (PS)	-2.7×10^{-5}	0.29×10^{-5}
PS x APC	-3.8×10^{-7}	1.1×10^{-7}
Percentage of local population which is		
i) Black and Asian	0.01	0.0010
ii) Armed Forces	0.31	0.05
iii) Institutions	0.03	0.0046
iv) Students	0.01	0.0037
v) Unemployed	0.01	0.0024
Population Estimate Level Variance	0.35	0.0091
Small Area Level Variance	0.11	0.0089

Table 5: Predicted absolute percentage inaccurate from the model estimated in Table 4, given specified values of each explanatory variable

		%		
Overall		**3.13**		
Absolute population change 1981-1991		Population Size		
		1,000	Mean	30,000
	0%	**3.80**	**2.49**	**1.21**
	Mean	**4.09**	**3.13**	**1.11**
	50%	**5.53**	**3.76**	
Method Used				
Local Census		**2.42**		
Other		**3.25**		
Percentage of Population				
i) Black and Asian	0%	**2.82**		
	95%	**6.99**		
ii) in Armed Forces	0%	**3.05**		
	5%	**17.42**		
iii) in Institutions	0%	**2.94**		
	30%	**8.44**		
iv) Students	0%	**3.01**		
	60%	**8.01**		
v) Unemployed	0%	**2.80**		
	50%	**4.87**		
Population Estimate Variation (95% CI)				
Lower Limit		**0.24**		
Upper Limit		**12.73**		
Small Area Variation (95% CI)				
Lower Limit		**0.89**		
Upper Limit		**8.04**		

Notes. (1) The value shown in bold type is in each case the inaccuracy predicted when all explanatory variables except the one shown in light type take their mean value. (2) The value shown in bold type is the predicted response exponentiated and with 1 subtracted, so that it may be interpreted directly as the predicted absolute percentage inaccuracy.

Figure 3: Absolute percentage inaccuracy of population estimates for 29 producers